Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

PROJECT REPORT
ON
FIA F1 1950-2020

Submitted by

Palagiri.Madhukar Reddy

11904176

Programme: Bachelor of Technology
Section: KM006
Course Code: INT217

Under the Guidance of
Komal Arora: 17783
Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science & Engineering
Lovely Professional University, Phagwara

DECLARATION

I, Palagiri.Madhukar Reddy, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work andis genuine.

Palagiri.Madhukar reddy
Registration No: 11904176
Signature

ACKNOWLEDGEMENT

Primarily I'd thank God for being able to complete my project with success. Then I'd like to thank my teacher **Ms. Komal Arora**, whose valuable guidance has been the ones that helped me patch this project and make it full proof success in contribution towards the completion of this project.

Last but not least I'd rather thanks to **Lovely Professional University,** and my parent's inspiration, who gave me this golden opportunity to learn many new things, to learn another aspect of life.

Palagiri.Madhukar reddy

CONTENTS:

Sr No.	Title	Page No.
1	Introduction	5
2	Objectives/Scope of the Analysis	6
3	Source of dataset	7 - 9
4	ETL Process	10
5	Power Pivot	11 - 12
6	Analysis of dataset	13 - 27
7	List of Analysis with results	28 - 35
8	Final Dashboard	36
9	Bibliography	37

INTRODUCTION

- Data management is important because the data your organization creates is a very valuable resource.
- The last thing you want to do is spend time and resources collecting data and business intelligence, only to lose or misplace that information.
- In that case, you would then have to spend time and resources again to get that same business intelligence you already had.
- And on that data analysis is carried out which show visualization of our problems in efficient way.
- Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision- making.
- This project is based on such data analysis on FIA 1950-2020
- This file contains two data sheets

OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, the aim of this project is to give answer of given objectives in easy way:

1. Most Number of laps
2. Top 10 Race Winners
3. Top 10 Drivers with Most Career Points
4. Most Occurred Venues
5. Top 10 Succesfull race finishers

SOURCE OF DATASET:

Source of the data sheet:

2020 https://www.kaggle.com/aadiltajani/fia-f1-19502019-data

The dataset is based on FIA F1(formula 1)1950-2020

The columns included in the dataset are given below:

- Year
  Gives the participation year of participant
- Position
  Gives the position achieved by participant
- Driver no
  Gives which sequence number of driver(participant)
- Venue
  Gives the race venue for drivers
- Name
  Gives the driver's name
- Nametag
  Gives the name tag of driver
- Team
  Gives the team name
- Laps
  Gives the laps completed by driver

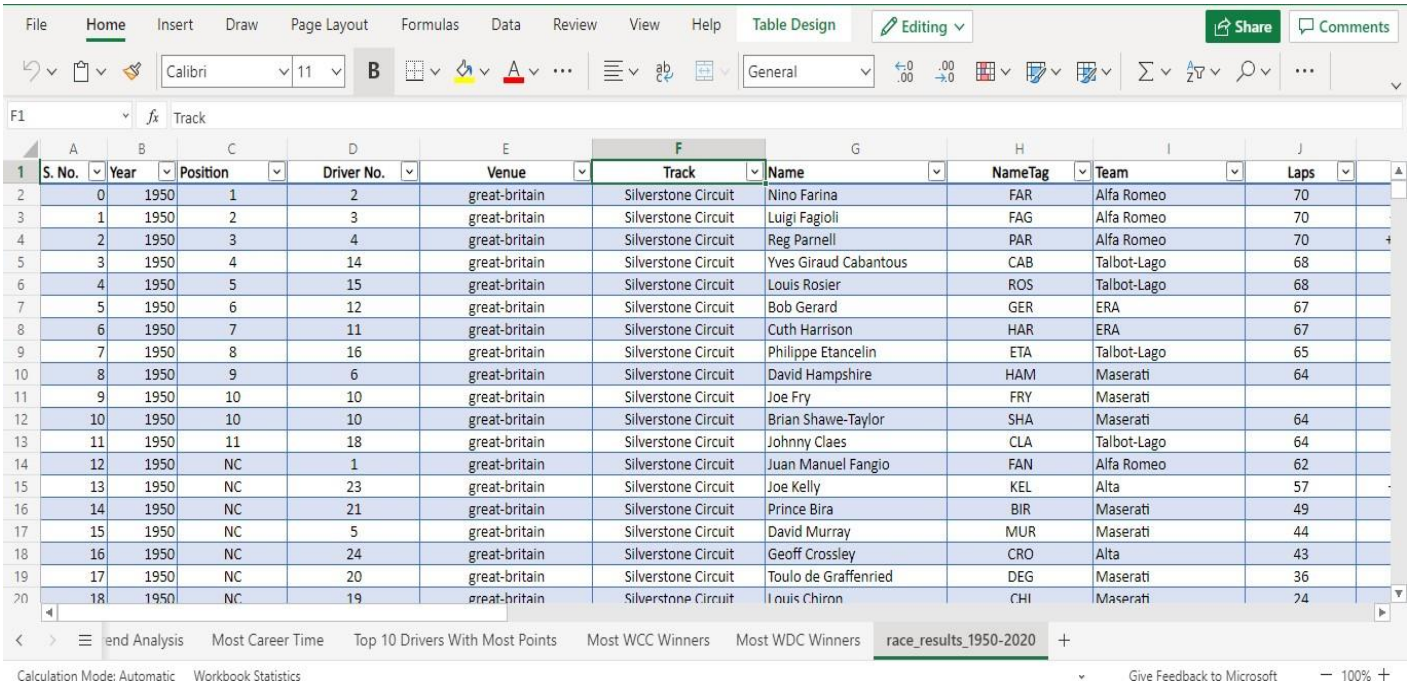
- Time

  Gives the time taken by driver to complete the race

ETL PROCESS:

- **ETL** is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.
- Full form of ETL is Extract, Transform and Load.
- The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:
  - Enhances Business Intelligence solutions for decision making.
  - Allows verification of data transformation, aggregation and calculations rules.
  - Allows sample data comparison between source and target system.
  - Helps to improve productivity as it codifies and reuses without additional technical skills.

Sample of dataset with data fields is given below:



S. No.	Year	Position	Driver No.	Venue	Track	Name	NameTag	Team	Laps
0	1950	1	2	great-britain	Silverstone Circuit	Nino Farina	FAR	Alfa Romeo	70
1	1950	2	3	great-britain	Silverstone Circuit	Luigi Fagioli	FAG	Alfa Romeo	70
2	1950	3	4	great-britain	Silverstone Circuit	Reg Parnell	PAR	Alfa Romeo	70
3	1950	4	14	great-britain	Silverstone Circuit	Yves Giraud Cabantous	CAB	Talbot-Lago	68
4	1950	5	15	great-britain	Silverstone Circuit	Louis Rosier	ROS	Talbot-Lago	68
5	1950	6	12	great-britain	Silverstone Circuit	Bob Gerard	GER	ERA	67
6	1950	7	11	great-britain	Silverstone Circuit	Cuth Harrison	HAR	ERA	67
7	1950	8	16	great-britain	Silverstone Circuit	Philippe Etancelin	ETA	Talbot-Lago	65
8	1950	9	6	great-britain	Silverstone Circuit	David Hampshire	HAM	Maserati	64
9	1950	10	10	great-britain	Silverstone Circuit	Joe Fry	FRY	Maserati	
10	1950	10	10	great-britain	Silverstone Circuit	Brian Shawe-Taylor	SHA	Maserati	64
11	1950	11	18	great-britain	Silverstone Circuit	Johnny Claes	CLA	Talbot-Lago	64
12	1950	NC	1	great-britain	Silverstone Circuit	Juan Manuel Fangio	FAN	Alfa Romeo	62
13	1950	NC	23	great-britain	Silverstone Circuit	Joe Kelly	KEL	Alta	57
14	1950	NC	21	great-britain	Silverstone Circuit	Prince Bira	BIR	Maserati	49
15	1950	NC	5	great-britain	Silverstone Circuit	David Murray	MUR	Maserati	44
16	1950	NC	24	great-britain	Silverstone Circuit	Geoff Crossley	CRO	Alta	43
17	1950	NC	20	great-britain	Silverstone Circuit	Toulo de Graffenried	DEG	Maserati	36
18	1950	NC	19	great-britain	Silverstone Circuit	Louis Chiron	CHI	Maserati	24

Power Pivot

- **Power Pivot** is a feature of Microsoft Excel. It is available as an add-in in Excel 2010, 2013 in separate downloads, and as an add-in included with the Excel 2016 program. Power Pivot extends a local instance of Microsoft Analysis Services Tabular that is embedded directly into an Excel Workbook.
- This allows a user to build a ROLAP model in Power Pivot and use pivot tables to explore the model once it is built.
- This allows Excel to act as a Self-Service BI platform, implementing professional expression languages to query the model and calculate advanced measures.

Analysis on dataset

1. Most Number of laps

- By performing this analysis, we will get the Comparison of drivers from year to year rather it was rapidly growing or decreasing and from that we can get Most number of laps with Races occured
- From this we can analyze reasons if the drivers decrease and can take certain actions for more participations

Description

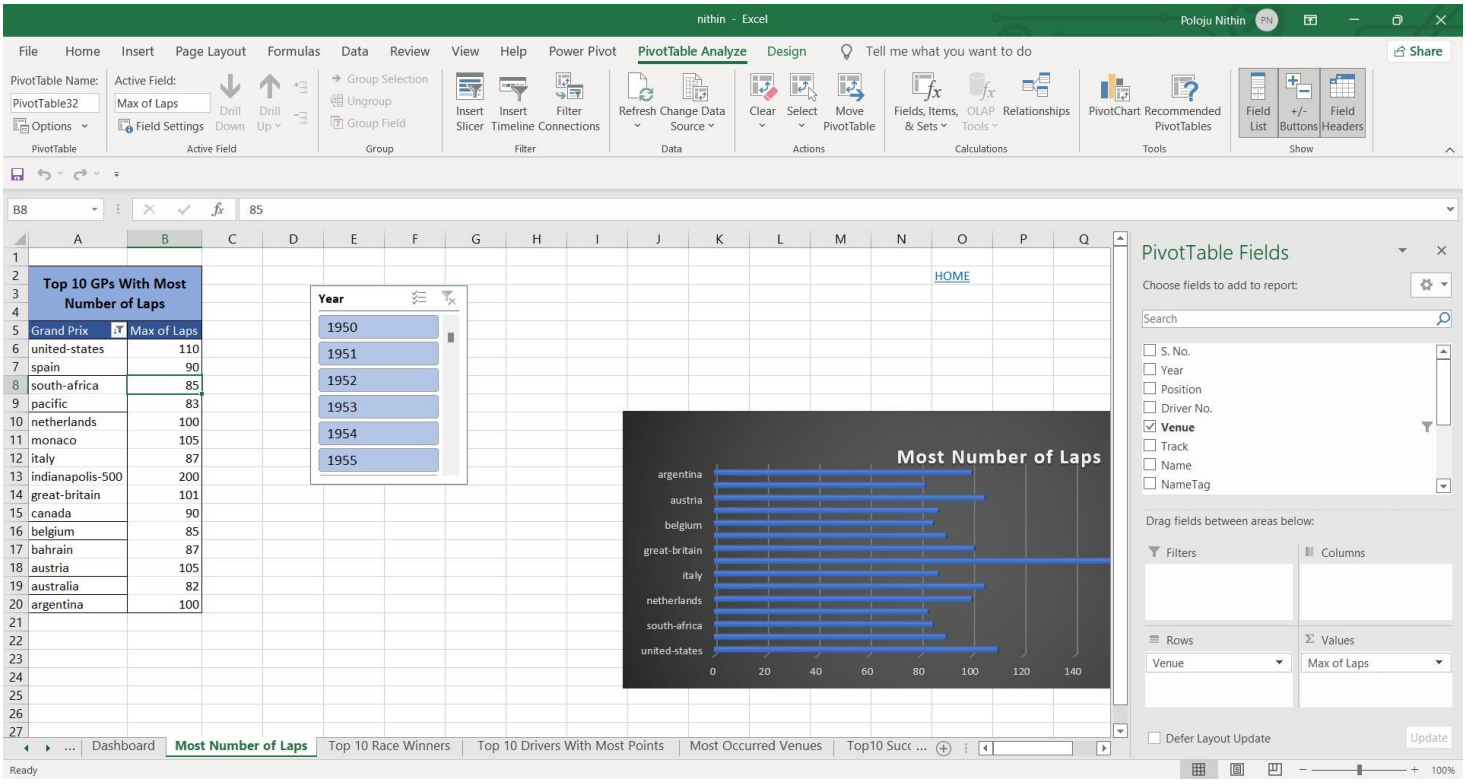
- The analysis is based on Laps completed by drivers, most wcc and wdc winners, top drivers with most career points, participations of drivers

Specific requirements, functions and formulas

- Some formulas are used for Comparisions.

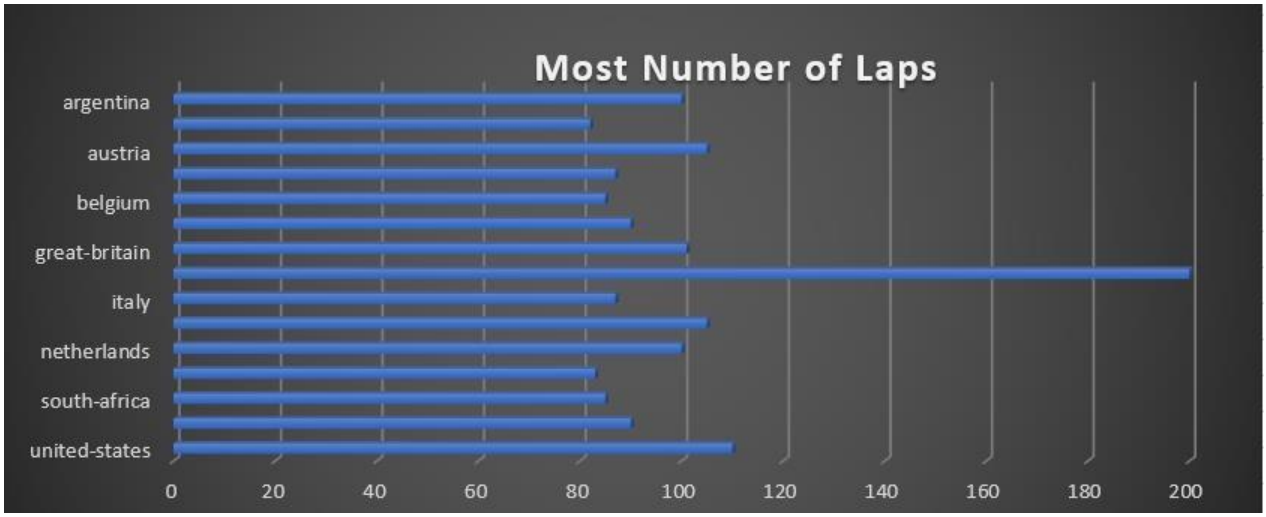
Analysis results

- When the participants are compared with respect to the number of laps


nithin - Excel
Poloju Nithin
File Home Insert Page Layout Formulas Data Review View Help Power Pivot PivotTable Analyze Design
Tell me what you want to do
Share
PivotTable Name: PivotTable32
Active Field: Max of Laps
Field Settings
Drill Down
Drill Up
Group Selection
Ungroup
Group Field
Insert Slicer
Insert Timeline
Filter Connections
Refresh
Change Data Source
Clear
Select
Move PivotTable
Fields, Items, & Sets
OLAP Tools
Relationships
PivotChart
Recommended PivotTables
Field List
+/- Buttons
Field Headers
PivotTable
Active Field
Group
Filter
Data
Actions
Calculations
Tools
Show
B8
85
Top 10 GPs With Most Number of Laps
Grand Prix
Max of Laps
united-states 110
spain 90
south-africa 85
pacific 83
netherlands 100
monaco 105
italy 87
indianapolis-500 200
great-britain 101
canada 90
belgium 85
bahrain 87
austria 105
australia 82
argentina 100
Year
1950
1951
1952
1953
1954
1955
HOME
Most Number of Laps
argentina
austria
belgium
great-britain
italy
netherlands
south-africa
united-states
0 20 40 60 80 100 120 140
PivotTable Fields
Choose fields to add to report:
Search
S. No.
Year
Position
Driver No.
Venue
Track
Name
NameTag
Drag fields between areas below:
Filters
Columns
Rows
Venue
Values
Max of Laps
Defer Layout Update
Update
Dashboard
Most Number of Laps
Top 10 Race Winners
Top 10 Drivers With Most Points
Most Occurred Venues
Top10 Succ ...
Ready
100%

Visualization

When the drivers are compared with number of laps



2.Top 10 Race winners

- By performing this analysis, we will get the Top 10 wcc winners

Description

- The analysis based on the Publishers of the game

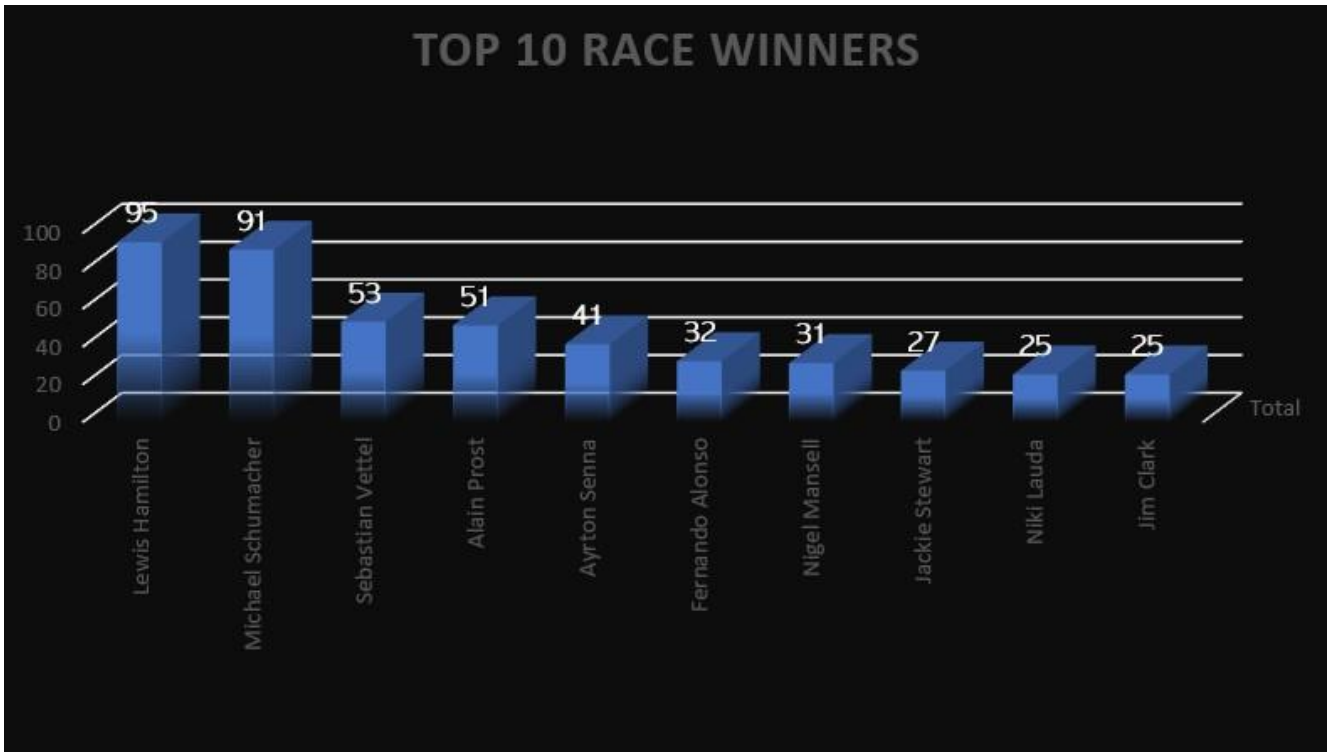
Specific requirements, functions and formulas

- count function is used in pivot table for the count of the wcc winners
- Sort function used to get the top 10 same as wdc winners

Analysis results

Drivers with Most Wins	
Driver	Wins
Lewis Hamilton	95
Michael Schumacher	91
Sebastian Vettel	53
Alain Prost	51
Ayrton Senna	41
Fernando Alonso	32
Nigel Mansell	31
Jackie Stewart	27
Niki Lauda	25
Jim Clark	25

Visualization



3.Top 10 drivers with most career points

Introduction

- By performing this analysis, we will get Topmost 10 drivers with highest points

Description

- The analysis based on the Genre of the race

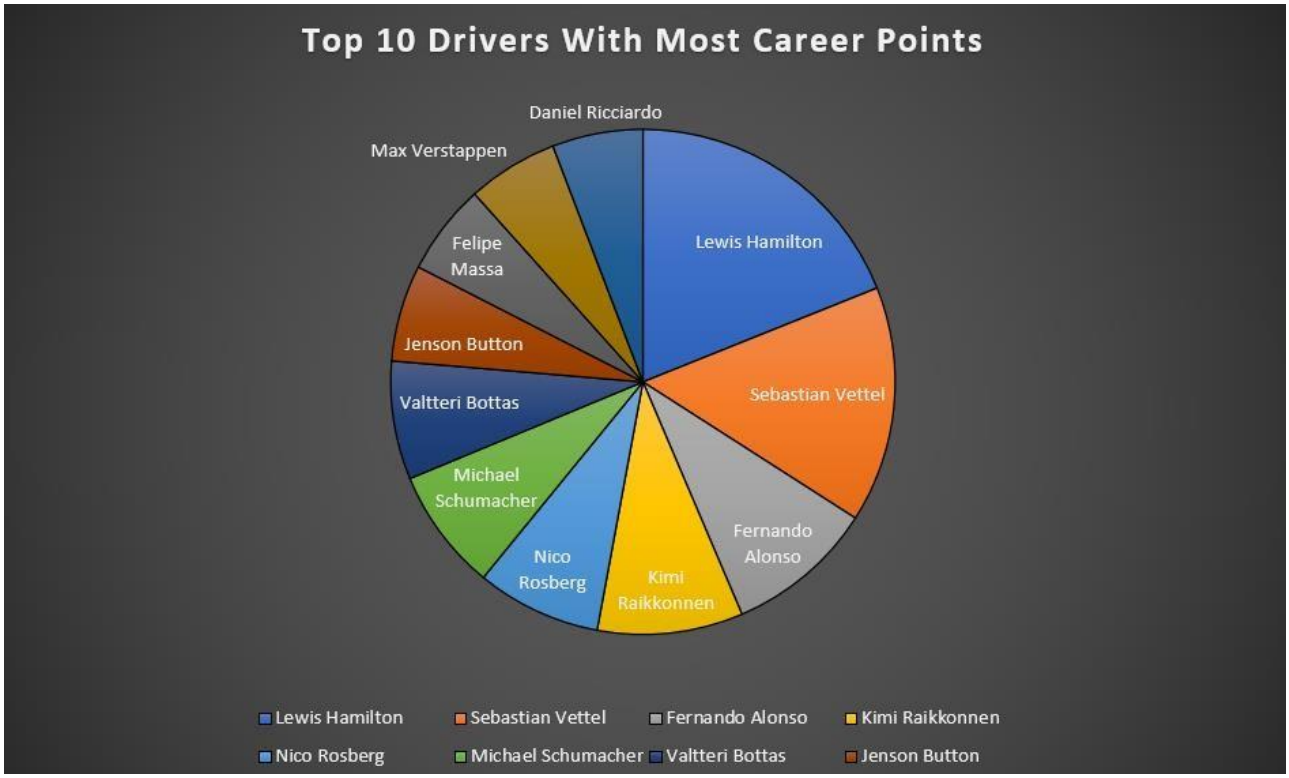
Specific requirements, functions, and formulas

- Count function is used in pivot table for the counting of no of points for a certain genre
- Sort function is used in pivot table to arrange in ascending order of the values of count of drivers.

Analysis results

	A	B
1		
2	Top Drivers With Most Career Points	
3		
4		
5	Driver	Career Points
6	Lewis Hamilton	3778
7	Sebastian Vettel	3018
8	Fernando Alonso	1899
9	Kimi Raikkonnen	1863
10	Nico Rosberg	1594.5
11	Michael Schumacher	1566
12	Valtteri Bottas	1512
13	Jenson Button	1235
14	Felipe Massa	1167
15	Max Verstappen	1162
16	Daniel Ricciardo	1159
17		
18		
19		
20		

Visualization



4.Most Occurred venues

Introduction

By performing this analysis, we will get Most occurred venues based on the races occurred places.

Description

- The analysis is based on Races and places

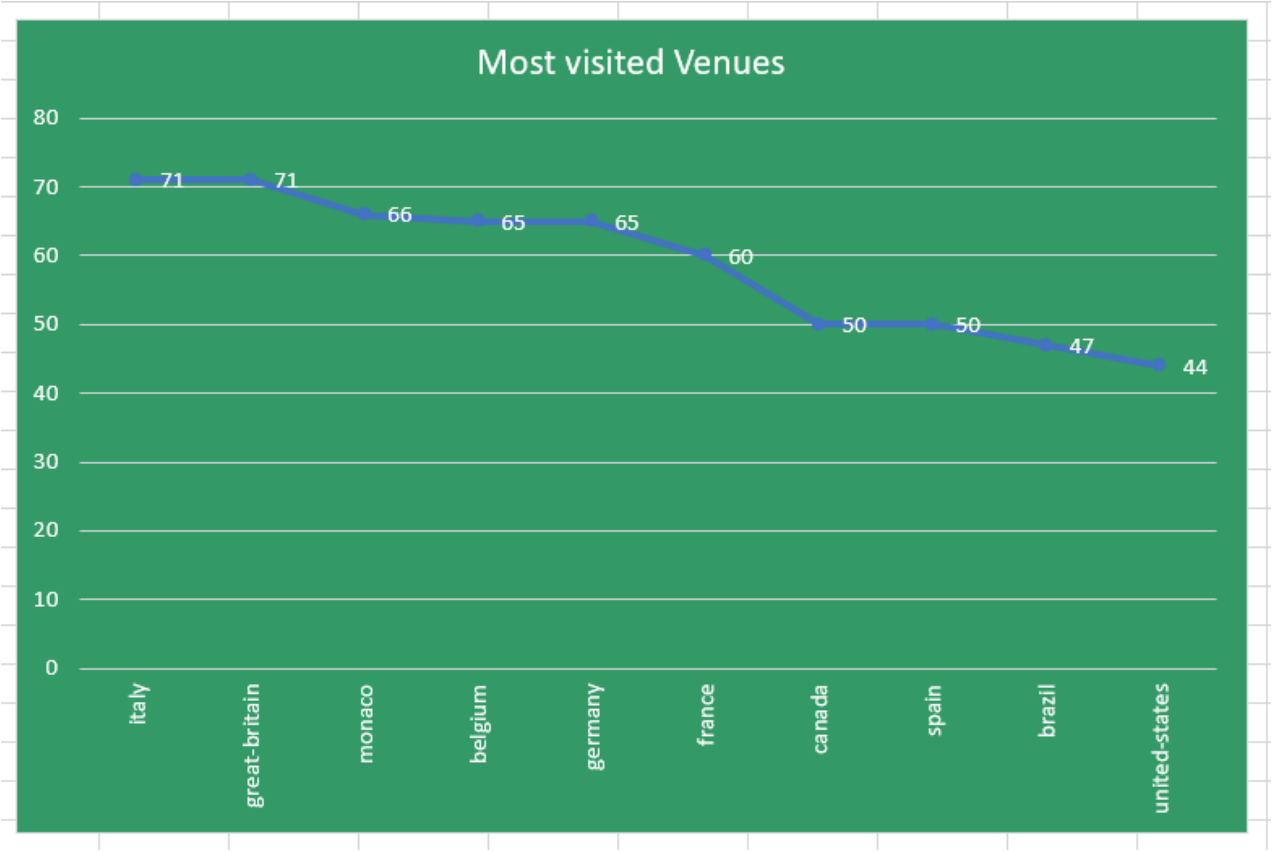
Specific requirements, functions, and formulas

- Value filters are used to for the top 10

Analysis results

Most visited venues by FIA Formula 1	
Venues	Race Count
italy	71
great-britain	71
monaco	66
belgium	65
germany	65
france	60
canada	50
spain	50
brazil	47
united-states	44

Visualization



5. Top 10 Most successful Race Finishers

Introduction

- By performing this analysis, we will get year to year comparisons in races and Successful race finishers from 1950-2020

Description

- The analysis is based on Names of the drivers.
- Drivers who DNF(do not finished races).

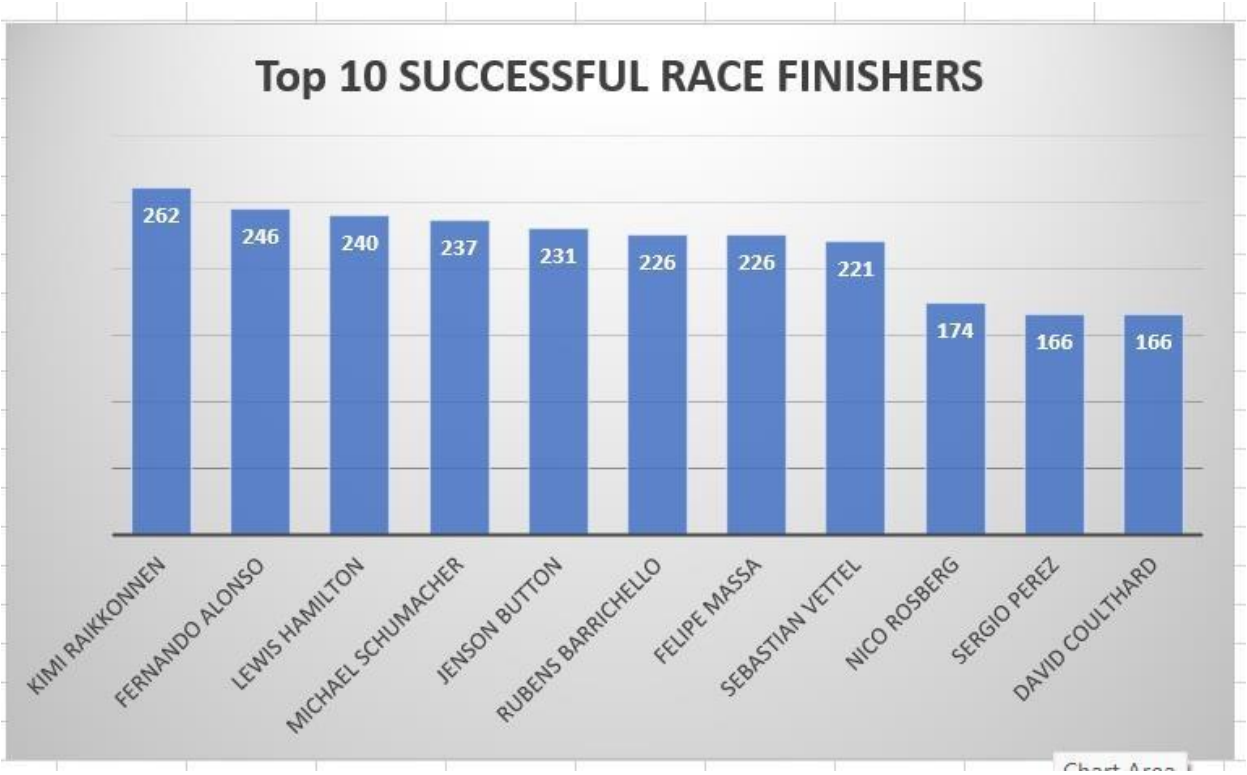
Specific requirements, functions and formulas

- Count function is used in pivot table for the Completed races and drivers participated
- Value Filter is used to show Names of the drivers

Analysis results

Drivers With Most Successful Race Finishes	
DNF	0
Row Labels	Successful Finishes
Kimi Raikkonnen	262
Fernando Alonso	246
Lewis Hamilton	240
Michael Schumacher	237
Jenson Button	231
Rubens Barrichello	226
Felipe Massa	226
Sebastian Vettel	221
Nico Rosberg	174
Sergio Perez	166
David Coulthard	166

Visualization



Conclusion

1. Most number of laps with venues and max of laps

Grand Prix

united-states

Spain

South-africa

Pacific

Netherlands

Monaco

Italy

Indianapolis-50

Great-Britain

Canada

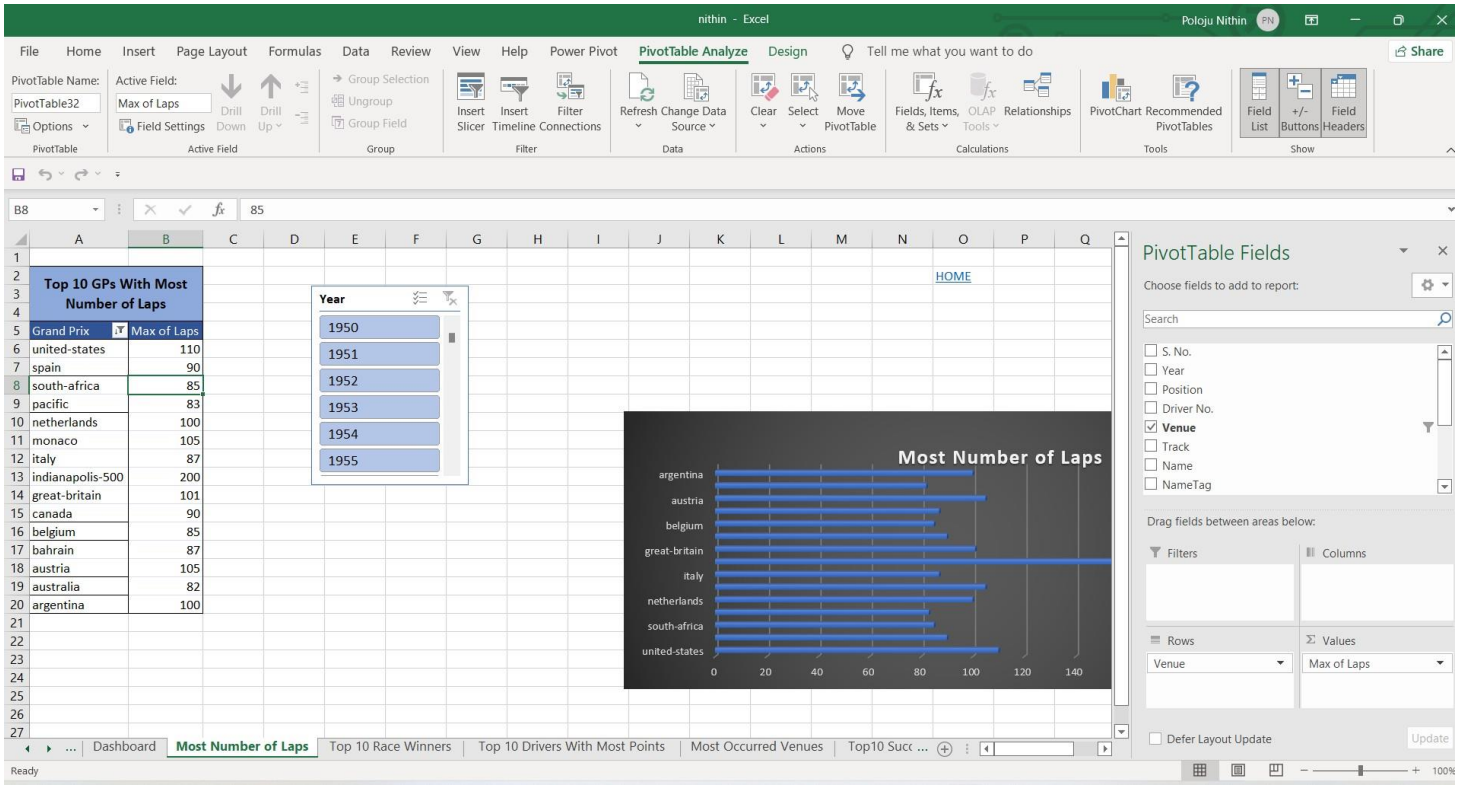
Belgium

Bahrain

Austria

Australia

Argentina


nithin - Excel
Poloju Nithin
File Home Insert Page Layout Formulas Data Review View Help Power Pivot PivotTable Analyze Design
Tell me what you want to do
Share
PivotTable Name: PivotTable32
Active Field: Max of Laps
Options
Field Settings
Drill Down
Drill Up
Group Selection
Ungroup
Group Field
Insert Slicer
Insert Timeline
Filter Connections
Refresh
Change Data Source
Clear
Select
Move PivotTable
Fields, Items, & Sets
OLAP Tools
Relationships
PivotChart
Recommended PivotTables
Field List
+/- Buttons
Field Headers
PivotTable
Active Field
Group
Filter
Data
Actions
Calculations
Tools
Show
B8
85
Top 10 GPs With Most Number of Laps
Grand Prix
Max of Laps
united-states 110
spain 90
south-africa 85
pacific 83
netherlands 100
monaco 105
italy 87
indianapolis-500 200
great-britain 101
canada 90
belgium 85
bahrain 87
austria 105
australia 82
argentina 100
Year
1950
1951
1952
1953
1954
1955
HOME
Most Number of Laps
argentina
austria
belgium
great-britain
italy
netherlands
south-africa
united-states
0 20 40 60 80 100 120 140
PivotTable Fields
Choose fields to add to report:
Search
S. No.
Year
Position
Driver No.
Venue
Track
Name
NameTag
Drag fields between areas below:
Filters
Columns
Rows
Venue
Values
Max of Laps
Defer Layout Update
Update
Dashboard
Most Number of Laps
Top 10 Race Winners
Top 10 Drivers With Most Points
Most Occurred Venues
Top10 Succ ...
Ready
100%

2.Top 10 Race winners

Lewis Hamilton

Michael Schumache

Sebastian Vettel

Alain Prost

Ayrton Senna

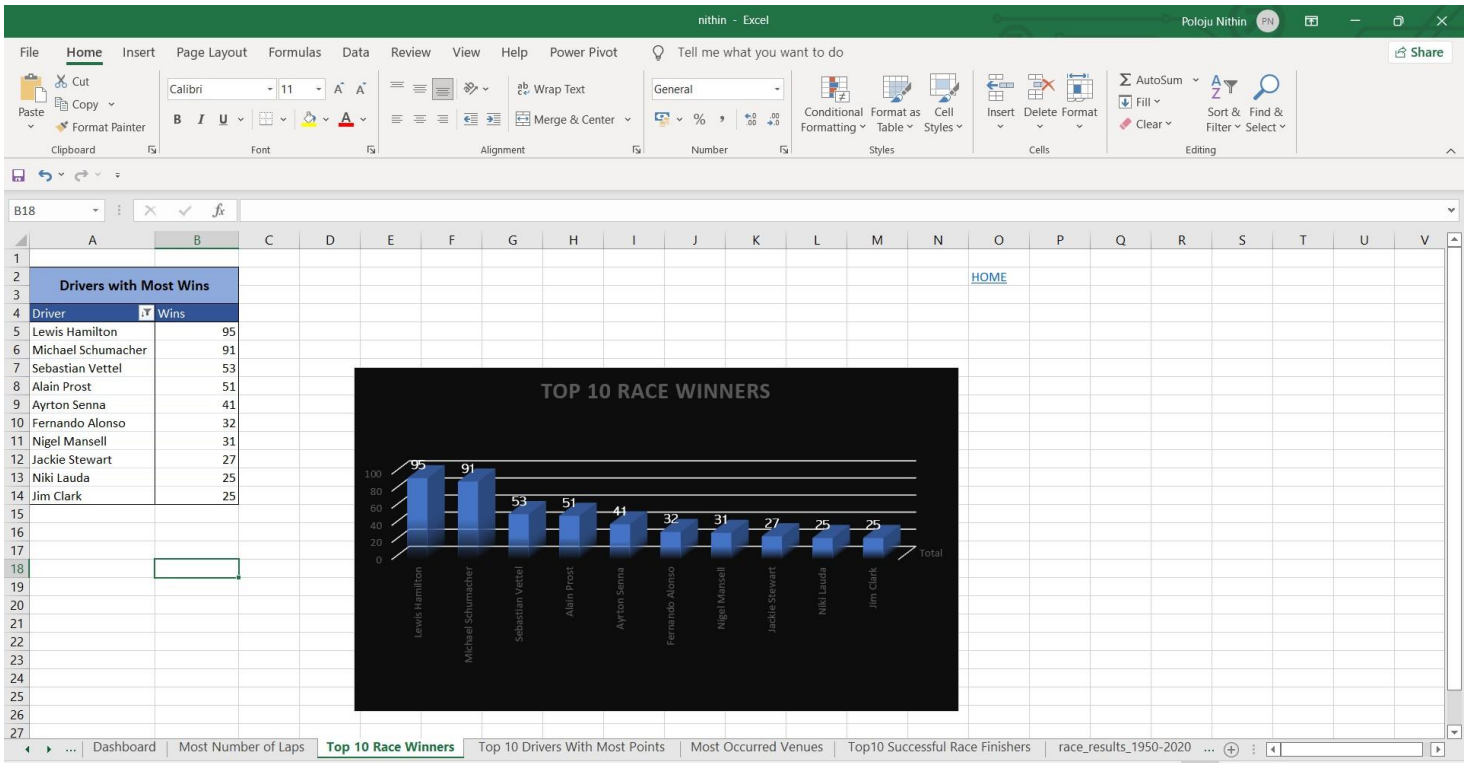
Fernando Alonso

Nigel Mansell

Jackie Stewart

Niki Lauda

Jim Clark



3.Top 10 drivers with most career points

Lewis Hamilton

Sebastian Vettel

Fernando Alonso

Kimi Raikkonnen

Nico Rosberg

Michael Schumacher

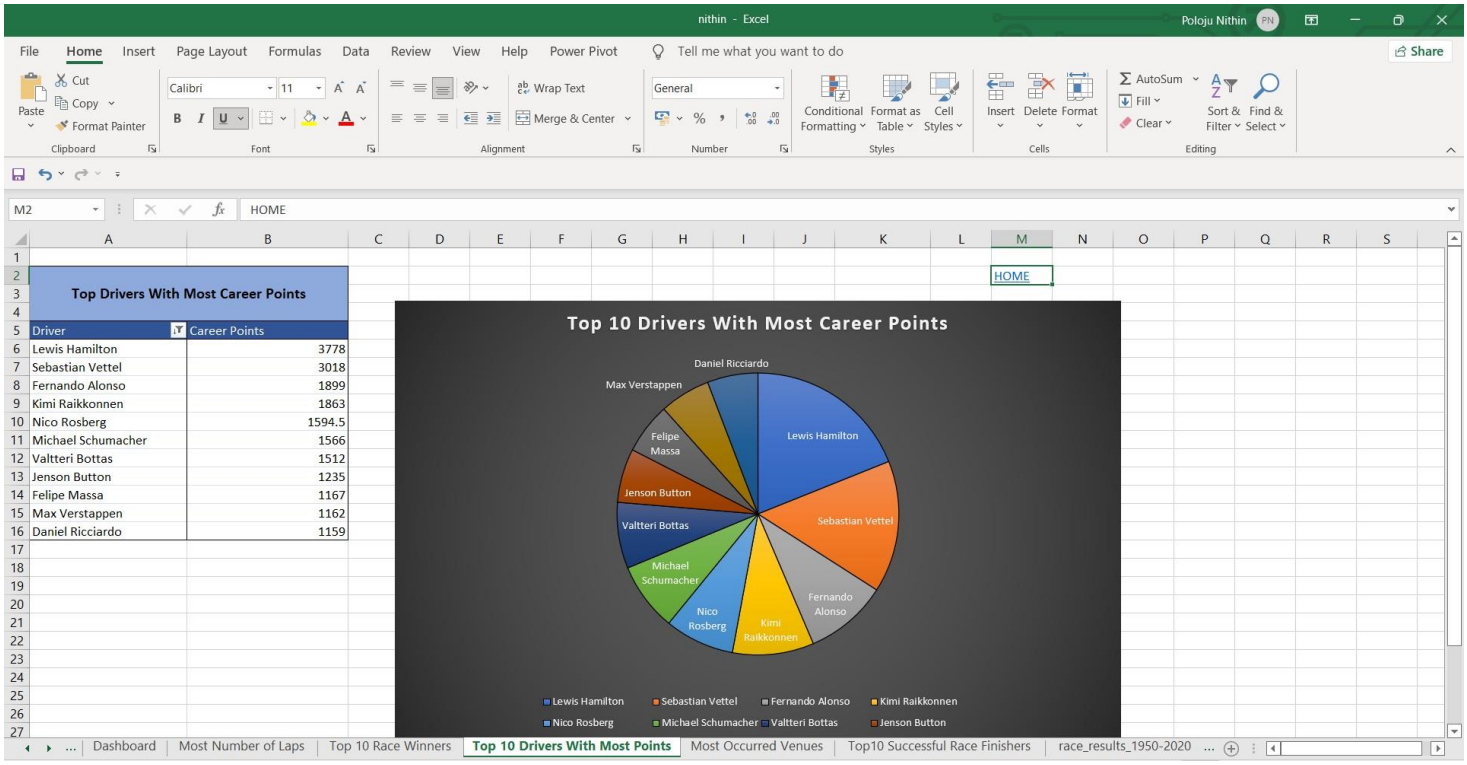
Valtteri Bottas

Jenson Button

Felipe Massa

Max Verstappen

Daniel Ricciardo


nithin - Excel
Poloju Nithin
File Home Insert Page Layout Formulas Data Review View Help Power Pivot
Tell me what you want to do
Share
Paste Cut Copy Format Painter
Calibri 11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert Delete Format
AutoSum Fill Clear
Sort & Filter
Find & Select
Clipboard Font Alignment Number Styles Cells Editing
M2
HOME
Top Drivers With Most Career Points
Driver
Career Points
Lewis Hamilton 3778
Sebastian Vettel 3018
Fernando Alonso 1899
Kimi Raikkonnen 1863
Nico Rosberg 1594.5
Michael Schumacher 1566
Valtteri Bottas 1512
Jenson Button 1235
Felipe Massa 1167
Max Verstappen 1162
Daniel Ricciardo 1159
HOME
Top 10 Drivers With Most Career Points
Daniel Ricciardo
Max Verstappen
Felipe Massa
Lewis Hamilton
Jenson Button
Sebastian Vettel
Valtteri Bottas
Michael Schumacher
Fernando Alonso
Nico Rosberg
Kimi Raikkonnen
Lewis Hamilton
Sebastian Vettel
Fernando Alonso
Kimi Raikkonnen
Nico Rosberg
Michael Schumacher
Valtteri Bottas
Jenson Button
Dashboard
Most Number of Laps
Top 10 Race Winners
Top 10 Drivers With Most Points
Most Occurred Venues
Top10 Successful Race Finishers
race_results_1950-2020

4. Most Occurred venues

italy

great-britain

monaco

belgium

germany

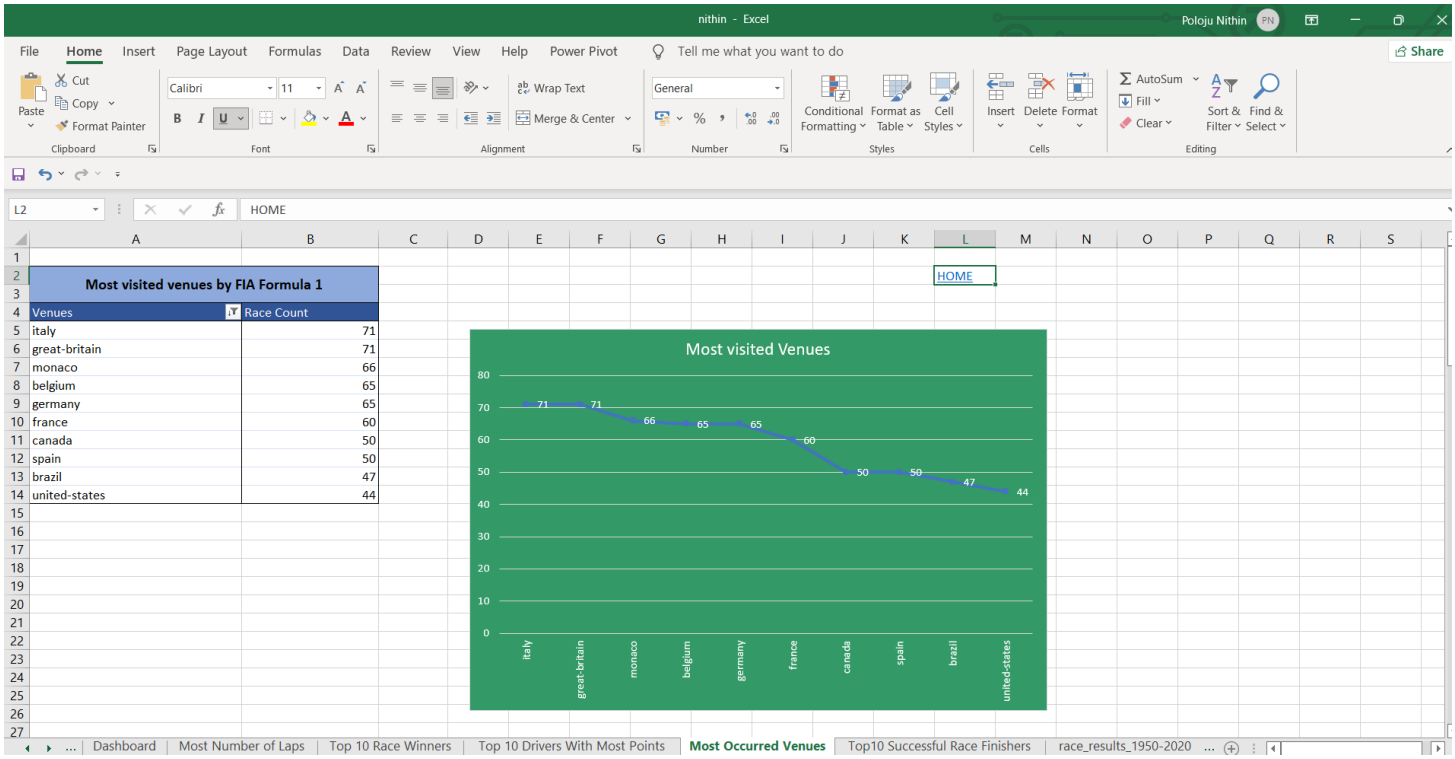
france

canada

spain

brazil

united-states



5.Top 10 succesful race finishers

Kimi Raikkonnen

Fernando Alonso

Lewis Hamilton

Michael Schumacher

Jenson Button

Rubens Barrichello

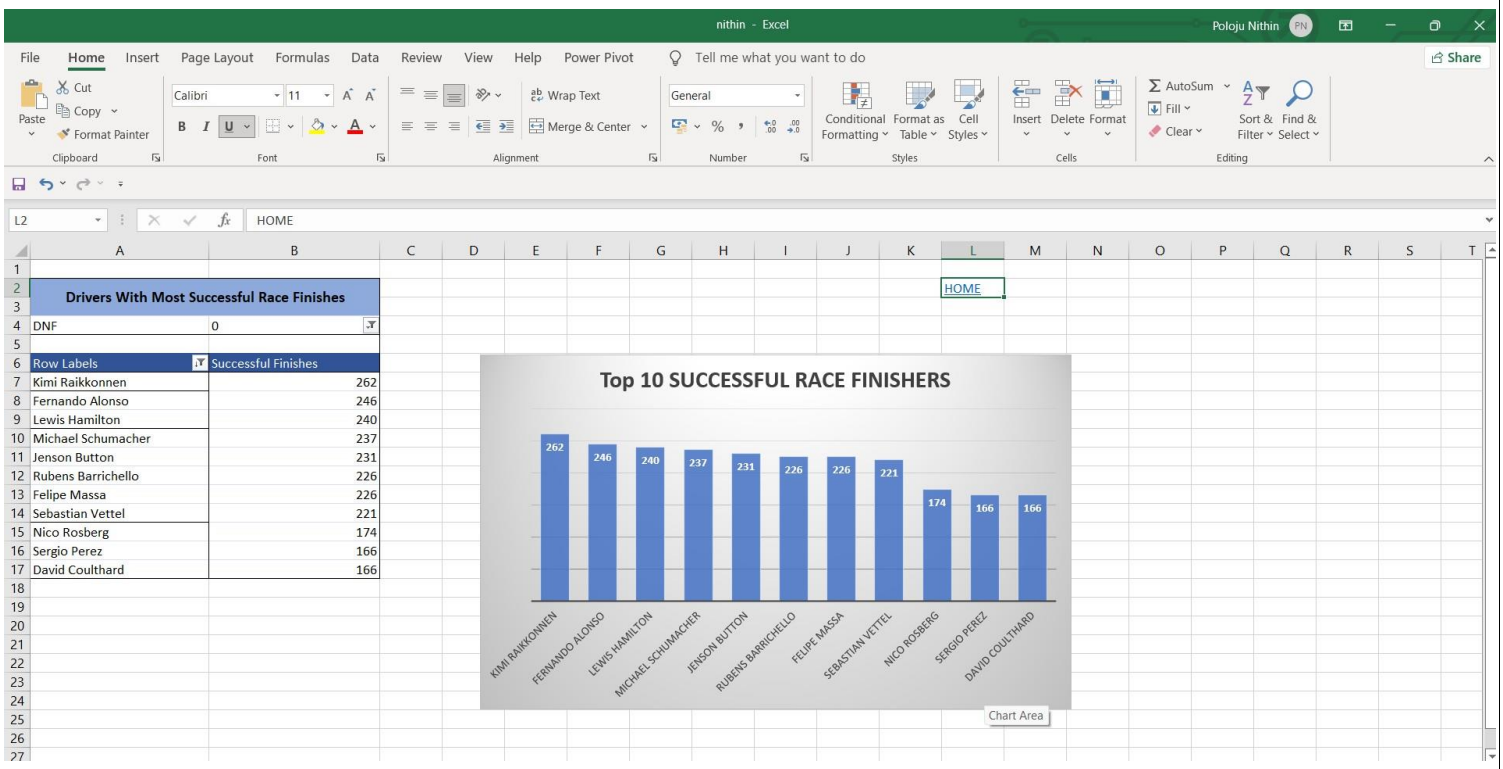
Felipe Massa

Sebastian Vettel

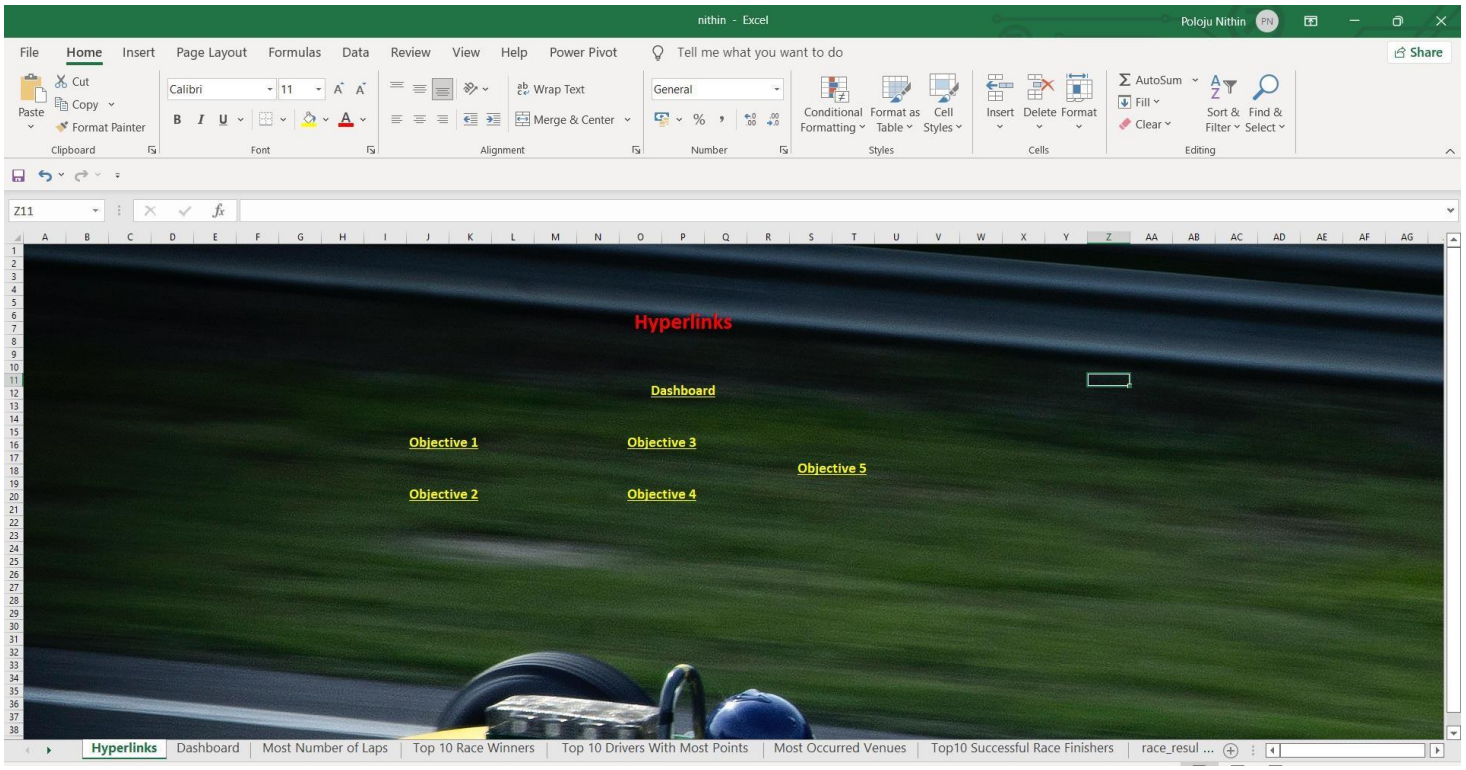
Nico Rosberg

Sergio Perez

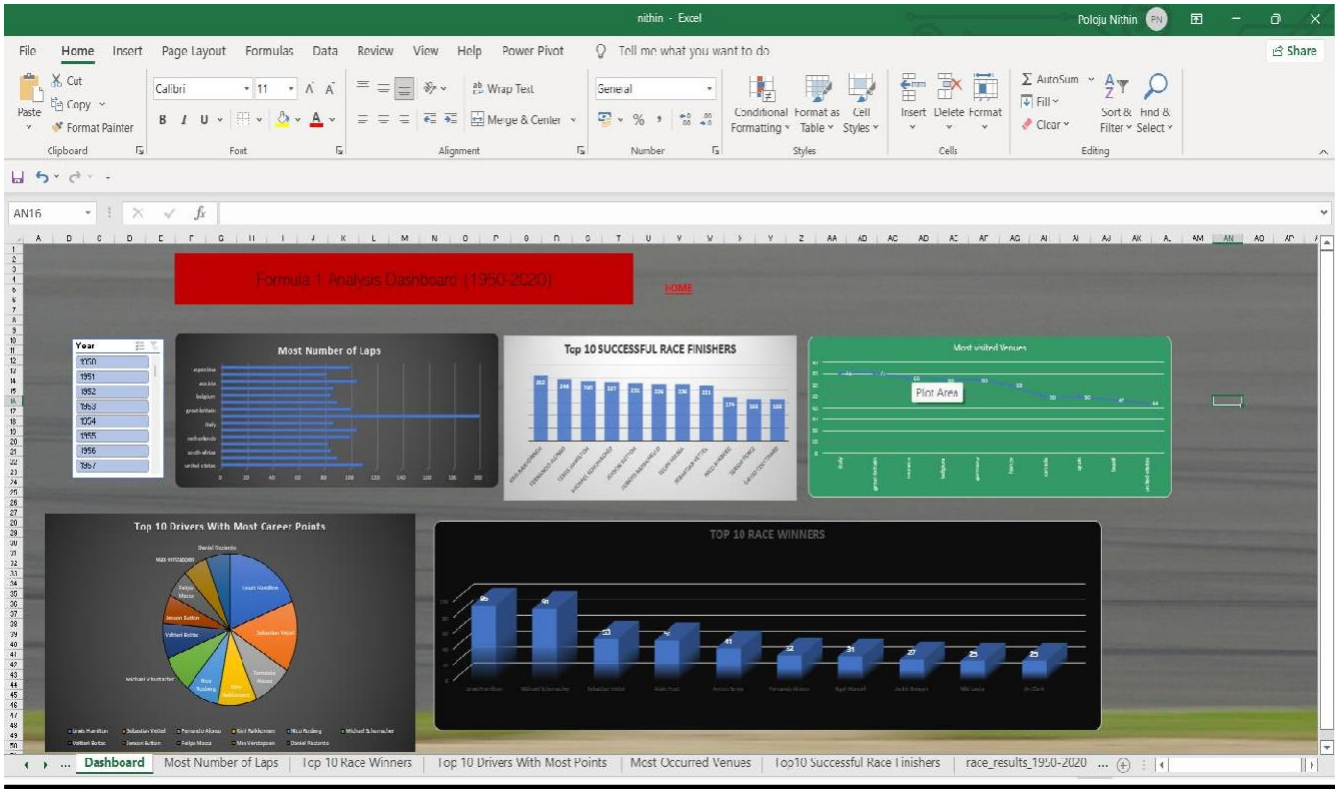
David Coulthard


nithin - Excel
Poloju Nithin
HOME
Drivers With Most Successful Race Finishes
DNF 0
Row Labels | Successful Finishes
Kimi Raikkonnen 262
Fernando Alonso 246
Lewis Hamilton 240
Michael Schumacher 237
Jenson Button 231
Rubens Barrichello 226
Felipe Massa 226
Sebastian Vettel 221
Nico Rosberg 174
Sergio Perez 166
David Coulthard 166
Top 10 SUCCESSFUL RACE FINISHERS
262
246
240
237
231
226
226
221
174
166
166
KIMI RAIKKONNEN
FERNANDO ALONSO
LEWIS HAMILTON
MICHAEL SCHUMACHER
JENSON BUTTON
RUBENS BARRICHELLO
FELIPE MASSA
SEBASTIAN VETTEL
NICO ROSBERG
SERGIO PEREZ
DAVID COULTHARD
Chart Area

Hyperlinks:


nithin - Excel
Poloju Nithin
Hyperlinks
Dashboard
Objective 1
Objective 3
Objective 5
Objective 2
Objective 4
Hyperlinks
Dashboard
Most Number of Laps
Top 10 Race Winners
Top 10 Drivers With Most Points
Most Occurred Venues
Top10 Successful Race Finishers

FINAL DASHBOARD:



BIBLIOGRAPHY:

- Dataset source:

https://www.kaggle.com/aadiltajani/fia-f1-19502019-data?select=race_results_1950-2020.csv

- Dashboard Background Image:

https://unsplash.com/photos

- Information about Data Management:

https://www.blue-pencil.ca/what-is-data-management-and-why-it-is- important/